



ŞANS BAŞARISININ TEST EŞİTLEMeye ETKİSİ

THE IMPACT OF CHANCE SUCCESS ON TEST EQUATING

Sibel BOZDAĞ*, Adnan KAN**

ÖZET: Bu araştırmanın amacı, şans başarısının test eşitlemeye etkisi olup olmadığını incelemektir. Araştırmada, tek grup düzeneği, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılmıştır. Verilerin analizi üç aşamada gerçekleştirilmiştir. İlk aşamada şanstın arındırılmış ve arındırılmamış test puanları için eşitleme koşullarının sağlanıp sağlanmadığı kontrol edilmiş, her iki durum için de testlerin ortalama güçlükleri, güvenilirlikleri, ortalamaları ve varyansları arasında 0,05 düzeyinde anlamlı bir fark olmadığı görülmüştür. Yapılan faktör analizi sonucunda her iki testin aynı yapıyı ölçtüğü ve tek faktör oldukları görülmüştür. İkinci aşamada, doğrusal ve eşit yüzdelikli eşitleme yöntemleri kullanılarak eşitlenmiş puanlar bulunmuştur. Üçüncü aşamada ise, bu yöntemlere ait ağırlıklandırılmış hata kareleri ortalamaları bulunmuş ve hata kareleri ortalaması en düşük olan yöntem en uygun eşitleme yöntemi olarak kabul edilmiştir. Araştırmanın sonucunda, şanstın arındırılmamış test puanlarının eşitlenmesinde doğrusal eşitlemenin, şanstın arındırılmış test puanlarının eşitlenmesinde ise eşit yüzdelikli eşitlemenin uygun yöntemler olduğu bulunmuştur. Tüm hatalar göz önünde bulundurulduğunda, şanstın arındırılmış eşit yüzdelikli eşitleme yöntemine ait hatanın en düşük olduğu görülmüştür.

Anahtar sözcükler: test eşitleme, doğrusal eşitleme, eşit yüzdelikli eşitleme, test eşitleme düzenekleri, şans başarısı

ABSTRACT: The objective of this study is to find out whether chance scores effects test equating, or not. Single group design, and linear equating and equipercentile equating were used in the study. The data analysis was performed at three stages. At the first stage, it was checked whether equating conditions are fulfilled for test score with and without the chance success. It was identified that there is no significant difference at 0.05 level between the two situations with respect to the mean difficulties, reliabilities, means and variances. Furthermore, at the end of the factor analysis performed, it can be said that both tests measure the same construct. At the second stage, the equated points were found by using linear and equipercentile equating. At the third stage, weighted mean square error of these methods was found, and the method with the lowest weighted mean square error was accepted as the most suitable equating method. At the end of the study, it was found out that; while linear equating is the most suitable method for equating of test points with chance success; equipercentile equating is the most suitable method for equating test points without chance success. Overall, it was observed in the study that the equating error is the lowest for equipercentile equating without chance success.

Keywords: test equating, linear equating, equipercentile equating, test equating design, chance scores

1. GİRİŞ

Test gizliliğinin korunamaması, ölçülen değişkenin doğasından gelen sorunlar, geçerlik ve güvenilirlik araştırmaları gibi nedenlerle testlerin birden fazla formları geliştirilir (Turgut, 1979). Bu formlarda kullanılan sorular benzer içeriğe ve istatistiksel özelliklere sahip olsalar da, madde güçlük düzeyleri arasında fark vardır. Aynı yeteneği, davranışı veya başarıyı ölçen testlerin farklı formlarından alınan puanların karşılaştırılması, eğitimde önemli yer tutar (Tsai, 1997). Aynı testin farklı formlarından alınan puanların karşılaştırılması için geliştirilen yöntem test eşitleme denir. Test eşitleme, aynı yapıyı ölçen iki formdan alınan puanlar arasında istatistiksel ilişki kurmak olarak tanımlanabilir (Felan, 2002). Böylece iki testten alınan puanlar aynı ölçek üzerinde yer alır ve karşılaştırılabilir.

Angoff (1971), test eşitlemeyi bir formun birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlar.

Test eşitlemenin iki amacı vardır (Barnard, 1996). Bunlar:

- Farklı test formlarını alan bireyler arasında yanlılığı önlemek
- Farklı formlardan alınan puanların aynı ölçek üzerinde rapor etmek ve rapor edilen puanların anlamını korumak.

* Öğretmen-Ölçme Değerlendirme uzmanı, sibelbozdog2002@yahoo.com

** Doç. Dr. Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, adnankan@gazi.edu.tr

Farklı formlardan elde edilen test puanlarının eşitlenmesi ile bireylerin gelişimini ölçmek, eğilimlerini belirlemek ve performanslarını karşılaştırmak mümkün olabilir. Test puanları eşitlenmeden önce bazı koşulların sağlanması gerekir. Bunlar:

I. Paralellik (Hambelton, 1985; Kelecioğlu, 1994, Şahhüseyinoğlu, 2005). Gulliksen (1967), iki farklı testten alınan gerçek puanların eşit olması durumunda, iki testin paralel olduğunu belirtir. Bu durumun gerçekte sağlanması zordur. Bu nedenle iki testin paralel olması için şu koşulların sağlanması gerekir.

- İki test aynı değişkeni ölçmelidir (Gulliksen, 1967; Angoff, 1971; Thorndike, 1982; Woldbeck, 1998; Tanguma, 2000; Felan, 2002). Woldbeck (1998), iki testin aynı zamanda tek boyutlu bir yapıya sahip olmaları gerektiğini belirtmiştir.

- İki testin güvenilirlikleri aynı olmalıdır (Angoff, 1971; Turgut, 1979; Crocker, Algina, 1986; Tanguma, 2000). Thorndike (1982) ve Angoff (1971) güvenilirliğin yüksek ve benzer olmaması durumunda formların anlamlı şekilde eşitlenmeyeceğini belirtmektedirler.

- İki testin ortalama güçlükleri aynı olmalıdır.

- İki testin ortalama ve varyansları eşit olmalıdır (Kelecioğlu, 1994; Şahhüseyinoğlu, 2005).

II. Simetriklik: İki formun birim sistemleri arasındaki dönüşüm, tesadüfi hatalardan kaynaklanan farklılıklar dışında tekdir. X formundan Y formuna dönüştürülen puanlar, Y formundan X formuna dönüştürülen puanlara eşittir (Angoff, 1971; Tanguma, 2000; Felan, 2002).

III. Gruptan Bağımsızlık: Eşitleme sonucunda elde edilen puanların, dönüştürmenin yapıldığı gruptan bağımsız olmasıdır (Angoff, 1971; Kelecioğlu, 1994 Felan, 2002). Eşitleme temel olarak hangi testin X, hangi testin Y olduğuna bakılmaksızın sonucun aynı olmasını gerektirir.

IV. Eşitlemenin başarılı şekilde gerçekleşebilmesi için formlar arasındaki korelasyonun yüksek olması gerekir (Dorans, 2000).

V. Formların madde sayıları eşit olmalı ve formlar aynı cevap formatına sahip olmalıdır (Masse, Allen, Wilson, Williams, 2006).

Eşitleme yapabilmek için göz önünde bulundurulması gereken iki önemli unsur vardır (Thorndike, 1982). Bunlar; eşitleme için verilerin hangi yöntemlerle toplandığı ve eşitleme için hangi istatistiksel yöntemlerin kullanıldığıdır.

1.1. Eşitleme Düzenekleri

Eşitleme yapmak için veri toplanması işlemine, 'eşitleme düzenekleri' adı verilir. Eşitleme düzenekleri, tarafsız ve ekonomik olmalıdır (Thorndike, 1982). Eşitleme düzenekleri hangi tür bilgilere ihtiyaç duyulduğu, iki puan dağılımının nasıl olduğu, grupların hangi formu alması gerektiği ve hangi yöntemin daha ekonomik olduğu göz önünde bulundurulur. Düzenek kurmanın 3 yolu vardır (Livingston, 2004). Bu yollar:

1. Bütün formlar aynı test grubuna uygulanır.
2. İki form, ölçülen özellik bakımından aynı olduğu düşünülen iki gruba uygulanır.
3. Gruplara, ölçülmek istenen özelliği ölçen, fakat bu formlardan farklı olan bir test (ankor) uygulanır.

Bu üç yol, iki formun eşitlenmesi için beş farklı düzenek kurulmasını gerektirir. Bu düzeneklerin hangisinin daha iyi sonuç verdiğini söylemek güçtür, her birinin bir diğerine göre avantajları ya da dezavantajları vardır. Eşitleme düzenekleri ilk aşamada, bu beş yöntemi kapsayacak biçimde ankor kullanılarak yapılan eşitleme ve ankor kullanılmadan yapılan eşitleme olarak ikiye ayrılır. Ankor kullanılmadan yapılan eşitleme; 'tek grup düzeneği', 'eşdeğer grup düzeneği' ve 'dengelenmiş grup düzeneği' olmak üzere üçe ayrılır. Bu düzenekler kullanıldığında eşitleme

ilişkisinin evrene genellenebileceği varsayılır (Livingston, 2004). Bu çalışmada eşitleme düzeneklerinden tek grup düzeneği kullanılmıştır. Tek grup düzeneğinde, iki veya daha fazla test aynı gruba uygulanır. Bu nedenle uygulaması oldukça basit eşitleme düzeneğidir. Felan'a (2002) göre, sadece bir gruba uygulandığı için eşitleme hatası diğer düzeneklerden daha küçüktür. Aynı grubun her iki formu da alması bu düzeneği güçlü kılan bir özelliktir ve diğer eşitleme düzenekleri ile karşılaştırıldığında eşitleme kesinliği daha yüksektir (Livingston, 2004). Livingston (2004), bu düzeneğin kullanılması durumunda, testi alan grubun hedef evreni temsil etmesine gerek olmadığını ve testi alan gruptan elde edilen eşitleme ilişkisinin hedef evrene genellenebileceğini belirtir.

Pratikte test eşitleme uygulamasında kullanılacak iki grubun seçkisiz olarak seçilmesi oldukça zordur (Tanguma, 2000). Farklı zamanlarda, farklı formları alan bireyler eşdeğer olmayabilir. İki grubun seçkisiz olarak seçilememesi durumunda, gruplar arasındaki farklılıkları kontrol edebilmek için her iki gruba uygulanan testlerden farklı sorulara sahip fakat aynı özelliği ölçen ortak test uygulanır. Bu ortak teste "ankor (anchor)" test denir. Ankor test gruplara her iki formdan farklı bir form olarak verilebileceği gibi testin bir parçası olarak da verilebilir. Bu durumda, her iki grup farklı iki formla birlikte ankor test alır. Eşitleme için en iyi ankor test, testlerin ölçtüğü niteliği ölçen ankor testtir. Ankor madde kullanılarak yapılan testlerde eşitlemenin amacına ulaşabilmesi için, ankor testin testlerle olan korelasyonunun yüksek ve iki formun uygulandığı gruplara ait puan dağılımlarının birbirine benzer olması gerekir. Testin ve soruların korunması amacıyla, aynı testin farklı formlarının kullanılması gereken durumlarda ankor düzenek önerilir (Tsai, 1997). Ankor düzenek, 'iç ankor' ve 'dış ankor' olmak üzere ikiye ayrılır.

İki formdan elde edilen puanları birbirine eşitlemek ya da iki puan dizisi arasında eşitlemeye olanak sağlayacak ilişkiyi belirlemek için istatistiksel yöntemler kullanılır. Bunlar;

1. Klasik Eşitleme Yöntemleri

- Doğrusal eşitleme
- Eşit yüzdelli eşitleme

2. Madde Tepki Kuramına (MTK) dayalı Eşitleme Yöntemleri olmak üzere iki kategoride incelenebilir. Bu çalışmada, klasik eşitleme yöntemleri kullanılmıştır.

1.2. Doğrusal Eşitleme

Standart puanlara dayanan eşitleme yöntemidir. İki testten elde edilen puanlar arasında ilişki kurmak için her iki testin ortalaması ve standart sapması kullanılır. Eşitlenmek istenen farklı iki test X ve Y olmak üzere, X formundan elde edilen ham puanlar ile Y formundan elde edilen ham puanlar arasında doğrusal bir ilişki olduğundan bu yöntem doğrusal eşitleme adını alır. Aynı standart puanlara karşılık gelen puanlar eşittir (Angoff, 1971).

X formundan Y formuna eşitlenmiş puanlar ile Y formundaki ham puanlar aynı ortalamaya ve standart sapmaya sahiptir. Bu durum (1) no'lu eşitlikte açıklanmıştır (Livingston, 2004):

$$\frac{Y - ortalama(Y)}{S_Y} = \frac{X - ortalama(X)}{S_X} \quad (1)$$

Bu eşitlik düzenlenirse,

$$Y^* = \left(\frac{S_Y}{S_X}\right)X + (ortalamaY - \left(\frac{S_Y}{S_X}\right).ortalamaX)$$

$$Y^* = \left(\frac{S_Y}{S_X}\right)X + (ortalamaY - \left(\frac{S_Y}{S_X}\right).ortalamaX) \quad (2)$$

Eğimi $\frac{S_y}{S_x}$, kesme noktası $(ortalamaY - (\frac{S_y}{S_x}).ortalamaX)$ olan doğru denklemi elde edilir.

Doğrusal eşitleme yöntemi, farklı test formlarını alan grupların aynı yetenek düzeyinde olması halinde uygulanır. Testin farklı formlarını alan gruplar, aynı yetenek düzeyinde değilse, farklı doğrusal eşitleme yöntemlerinin uygulanması önerilir

İki testin puan dağılımları aynı ve güvenilirlikleri eşit ise doğrusal eşitleme yönteminin kullanılması uygundur (Thorndike, 1982; Skaggs and Lissitz; 1986; Crocker, Algina; 1986; Woldbeck, 1998). Doğrusal eşitleme uygulaması kolay ve pratik bir yöntemdir.

1.3. Eşit Yüzdellikli Eşitleme

Doğrusal eşitleme için, eşitlenmek istenen iki formdan elde edilen puanların aynı dağılıma sahip olması gerektiği üzerinde durulur. İki form aynı puan dağılımına sahip değil ise eşit yüzdellikli eşitleme yöntemi önerilir (Zhu, 1998). X ve Y formlarından (X ve Y aynı güvenilirliğe sahip olmak üzere) elde edilen puanlar, aynı düzeltilmiş yüzdellik sıraya sahipse eşittir (Angoff, 1971). Bu eşitleme yönteminde Form X'in dağılımı, Form Y'nin dağılımına iki testin yüzdellik sıraları hesaplanarak eşitlenir. İki testin puan dağılımlarının farklı olması durumunda, eşitleme kesinliği, eşit yüzdellikli eşitleme ile sağlanır (Angoff, 1971; Thorndike, 1982). Eşit yüzdellikli eşitleme, doğrusal eşitlemeye kıyasla daha az sayılıya sahiptir (Woldbeck, 1998).

Eşit yüzdellikli eşitleme, iki basamakta gerçekleştirilir. İlk olarak her iki dağılıma ait puanların yüzdellik sıraları bulunur. Bu yüzdellik sıralarla ham puanlara ait grafik çizilir. İkinci basamakta ise, bu grafikten yararlanılarak elle veya analitik yöntemler ile eşitlenmiş puanlar bulunur.

Eşit yüzdellikli eşitleme için puan dağılımının gergin ve sıkışık (puan aralıklarının daha küçük) olması gereklidir, böylece bir formun puan dağılımı, diğer formun puan dağılımına denk gelir (Angoff, 1971; Woldbeck, 1998). Eşit yüzdellikli eşitleme yönteminin ilk basamağında, X ve Y formundan alınan puanların yüzdellik sıraları bulunur. Bu durum Şekil 2'deki diyagram üzerinde gösterilebilir.



Şekil 1: Eşit yüzdellikli eşitleme

Şekil 1'de görüldüğü üzere birinci diyagram yüzdellik sıraları gösterirken, 2. diyagram düzeltilmiş yüzdellik sıraları gösterir. Eşit yüzdellikli eşitleme yönteminin uygulanması karmaşık bir yapıya sahiptir. X formundaki her puan bazen Y formundaki yüzdellik sırasına karşılık gelmeyebilir. Bu nedenle çeşitli öteleme ve düzeltme çalışmaları yapılmaktadır. Düzeltme çalışmaları eşitlemeden önce yapılıyorsa “ön düzeltme (presmoothing)”, eşitleme işleminden sonra yapılıyorsa “son düzeltme (postsmoothing)” adı verilir. X formundaki puanların yüzdellik sıraları, Y formundaki yüzdellik sıralarına karşılık gelmediğinde öteleme yöntemleri kullanılarak eşitlenmiş puanlar hesaplanır

Livingston (2004) tarafından önerilen ve (3) no'lu eşitlikte verilen öteleme formülü bu çalışmaya uyarlanarak kullanılmıştır.

$$2003 \text{ puan} + \frac{2005 \text{ yüzdelik} - 2003 \text{ altyüzdelik}}{2003 \text{ üstyüzdelik} - 2003 \text{ altyüzdelik}} \quad (3)$$

Eşitleme yöntemlerinin uygunluğu, eşitleme hatasına dayalı olarak belirlenir. Eşitleme alanında yapılan birçok çalışmada çeşitli eşitleme hataları hesaplanmıştır. Bu hataların farklılığı, kavramsal çatılarının farklı olmasından kaynaklanır. Bunlardan biri de ağırlıklandırılmış hata kareleri ortalamasıdır. Eşitleme yöntemlerinden elde edilmiş puanların hata miktarını belirlemek için her bir ham puan ve bu ham puana karşılık gelen eşitlenmiş puanlar ağırlıklandırılmış hata kareleri ortalaması ile karşılaştırılır (Skaggs ve Lissitz, 1986; Kelecioğlu, 1994; Şahhüseyinoğlu, 2005).

$$WMSE = \sum_{i=1}^{k-1} f_i (X_E - X_{crit})^2 / \sum_{i=1}^k f_i S^2_Y \quad (4)$$

k : Y testindeki madde sayısı.

S^2_Y : Y testindeki ham puanların varyansı.

X_{crit} : Y testindeki i. ham puan.

X_E : Eşitleme yöntemleriyle elde edilen ve X testindeki i. ham puana eşit olan puan.

f_i : Y testindeki i. ham puan frekansı.

Bu çalışmada, şans başarısının test eşitlemeye etkisi incelenmiştir Şans başarısı, yanıtlayıcının çoktan seçmeli testte doğru cevabı yoklanmak istenen bilgiyle değil de, tahmine giderek bulmasıdır (Turgut, 1971). Şans başarısından doğan hata madde ve test istatistiklerine etki etmekte ve testin geçerliği ve güvenilirliğini düşürmektedir (Baykul, 2000). Bu çalışmada ham puanları, şans başarısından arındırmak için düzeltme formülü kullanılmıştır. En çok kullanılan düzeltme formülü ise yanlış cevapların bir kısmının doğru cevaplardan çıkarıldığı formüldür. Bu amaçla düzeltilmiş test puanı (D.T.P) bulunur (Turgut, 1971).

$$D.T.P = D - \frac{Y}{k-1} \quad (5)$$

D: Bireyin toplam doğru cevap sayısı.

Y: Bireyin yanlış cevap sayısı.

k: Seçenek sayısı.

1.4. Araştırmanın Amacı ve Önemi

Bu araştırmanın amacı, şans başarısının test eşitlemeye etkisi olup olmadığını incelemektir. Şanstan arındırılmış veya arındırılmamış ham puanların eşitlenmesinde kullanılan yöntemlerden hangisinin daha az eşitleme hatası içerdiğini bulmak amaçlanmıştır. Bir diğer ifade ile ham puanların, şans başarısından arındırılmadan ve şans başarısından arındırılarak eşitlenmesinde doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemlerinden hangisinin daha uygun olduğudur. Yapılan literatür taramasında, şans başarısının test eşitlemeye etkisini inceleyen bir çalışmaya rastlanamamıştır. Bu nedenle bu çalışmanın, test eşitleme alanında yapılan çalışmalara bir katkısı olacağı düşünülmektedir.

Araştırmanın bir diğer amacı ise, ham puanlar ile eşitlenmiş puanlar arasında bir eşitleme bağlantısı kurmaktır. Bu bağlantı doğrultusunda, birbirinin yerine kullanılabilen puanlar bulunmak istenmektedir. Ülkemizde farklı zamanlarda yapılan ve birkaç yıl geçerliliği olan KPSS, Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES), Kamu Personeli Yabancı Dil Bilgisi Seviye

Tespit Sınavı (KPDS) gibi sınavlar uygulanmaktadır. Bu sınavlardan farklı yıllarda alınan puanların aynı ölçek üzerinde yer almadığı sürece eşdeğer kabul edilmesi ve birbirinin yerine kullanılması doğru değildir. Puanların aynı ölçek üzerinde yer alması ise, test eşitleme ile sağlanır. Ülkemizde kısıtlı çalışma alanı bulan test eşitleme, bu noktada öne çıkmaktadır. Çalışmanın bir diğer amacı, bu noktaya dikkat çekmektir.

1.5. Alt Problemler

1.5a. Şans başarısından arındırılmamış 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, eşit yüzdelliği eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

1.5b. Şans başarısından arındırılmamış 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, doğrusal eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

1.5c. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılmadan eşitlenmesinde eşit yüzdelliği eşitleme ve doğrusal eşitleme yöntemlerinden hangisi daha uygundur?

1.5d. Şans başarısından arındırılan 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, eşit yüzdelliği eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

1.5e. Şans başarısından arındırılan 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, doğrusal eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanları nasıldır?

1.5f. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılarak eşitlenmesinde eşit yüzdelliği eşitleme ve doğrusal eşitleme yöntemlerinden hangisi daha uygundur?

1.5g. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılarak ve arındırılmayarak eşitlenmesinde kullanılan doğrusal eşitleme yöntemlerinden hangisi daha uygundur?

1.5h. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılarak ve arındırılmayarak eşitlenmesinde kullanılan eşit yüzdelliği eşitleme yöntemlerinden hangisi daha uygundur?

2. YÖNTEM

2.1. Araştırma Grubu

Araştırma, Mersin ilinde çeşitli ilköğretim okullarında öğrenim gören 1031 8. sınıf öğrencisi üzerinde yürütülmüştür.

2.2. Veri Toplama Aracı

Araştırma test eşitleme yöntemlerini karşılaştırmak, şans başarısının test eşitleme üzerindeki etkisini belirlemek amacıyla 2003 ve 2005 OKS Türkçe alt testleri ve bu testlerden elde edilen puanlar üzerinde yürütülmüştür.

Araştırmada kullanılan 2003 ve 2005 OKS Türkçe alt testlerine ait maddeler birleştirilerek 50 maddelik bir test oluşturulmuştur. Oluşturulan 50 maddelik bu test, 2005–2006 eğitim-öğretim yılı içerisinde yer alan çeşitli ilköğretim okullarında öğrenim gören 1031 8. sınıf öğrencisine uygulanmıştır. Birleştirilen test maddeleri, öğrencilere tek oturumda verilmiş, 50 dakikalık süre tanınmıştır. 2003 OKS Türkçe alt test puanları, 2005 OKS Türkçe alt test puanlarına, tek grup düzeneği kullanılarak eşitlenmiştir.

2.3. Verilerin Analizi

Bu çalışmada 2003 ve 2005 OKS Türkçe alt testlerine ait puanlar, şans başarısından arındırılmadan ve arındırılarak, iki farklı yöntem ile eşitlenmek istenmektedir. Eşitleme işlemine geçilmeden önce, her iki teste ait puanlar (5) no'lu eşitlikte verilen düzeltme formülüyle şanstı

arındırılmış ve iki farklı ham puan türü elde edilmiştir. Güçlük düzeyleri aynı olduğu düşünülen bu testler, benzer yetenek düzeyine sahip bireylere uygulandığı için yatay eşitleme yapılmıştır.

Verilerin analizi üç aşamada gerçekleştirilmiştir. Birinci aşamada, eşitleme koşullarının sağlanıp sağlanmadığı test edilmiştir. İkinci aşamada, eşitleme yöntemleri kullanılarak, eşitlenmiş puanlar elde edilmiştir. Üçüncü aşamada ise, her bir eşitleme yöntemine ait hata kareleri ortalamaları hesaplanmıştır.

I. Aşama: Verilerin analizinin birinci aşamasında, eşitleme koşullarının sağlanıp sağlanmadığı test edilmiştir

İlk olarak eşitlenmek istenen testlerin tek boyutlu olup olmadıkları ve aynı yapıyı ölçüp ölçmedikleri belirlenmeye çalışılmıştır. Eşitlenmek istenen iki testin de aynı yapıyı ölçüp ölçmediğine, her iki teste ait birey-madde matrisi birleştirilerek, temel bileşenler analizi ile bakılmıştır. Temel bileşenler analizi testlerden elde edilen madde puanlarının iki kategorili (dichotomous) olması sebebiyle tetrakorik korelasyon matrisi üzerinden yürütülmüştür ve bu analize ilişkin sonuçlar Tablo 1’de verilmiştir.

Tablo 1: 2003 ve 2005 OKS Testlerine ait Temel Bileşenler Analizi Sonuçları

Bileşen	2003 OKS		2005 OKS		BİRLEŞİK (2003 VE 2005 OKS)	
	Özdeğer	V.A.O (%)	Özdeğer	V.A.O (%)	Özdeğer	V.A.O (%)
1	8,46	33,82	8,45	33,80	15,53	31,07
2	1,30	5,25	1,67	6,70	2,05	4,12

Tablo 1’deki verilere göre gerek 2003 ve 2005 gerekse bu iki testi birleştirmek suretiyle elde edilmiş OKS test sonuçları üzerinde yürütülen temel bileşenler analizi sonucunda her iki teste ve birleştirilmiş teste ilişkin özdeğerler ve varyans açıklama oranları (V.A.O) birbirine çok yakın ve benzer değerlere sahiptir. Bu bulgu testlerin aynı yapıyı ve özelliği ölçtüğüne ilişkin kanıt olarak kullanılabilir. Aynı şekilde bu özdeğerler ve varyans açıklama oranları 1. faktörden sonra keskin bir düşüş göstermekte ve ilk faktör ile kendisinden sonra gelen ilk faktöre (2. faktör) ait V.A.O ile özdeğerler arasında en az 5-6 kat fark olduğu göze çarpmaktadır. Bu bulgu testlerin baskın tek faktöre sahip olduğuna ilişkin kanıt olarak kullanılabilir. Hambleton ve Swaminathan (1985) bir çok durumda tek boyutluluk varsayımının karşılanabilmesi için testin baskın tek bir faktöre sahip olmasını yeterli görmekte ve bu baskın faktörü testle ölçülmek istenen özellik (trait) olarak tanımlamaktadır.

Şanstan arındırılmamış 2003 ve 2005 OKS Türkçe alt testlerine ait bazı betimsel istatistikler hesaplanmış ve Tablo-2’de verilmiştir. Şanstan arındırılmamış test puanları arasındaki korelasyon 0,77 olarak bulunmuştur. Bu ilişkinin oldukça yüksek olduğu söylenebilir.

Tablo 2: Şans Başarısından Arındırılmamış 2003 ve 2005 OKS Türkçe Alt Test Puanlarına Ait Bazı Betimsel İstatistikler

Alt Testler	K	N	\bar{X}	S^2_x	S_x	KR – 20	Ortalama Güçlük	Çarpıklık	Basıklık
Türkçe 2003	25	1031	13.47	32,7	5,72	0,853	0,539	0,051	-1,063
Türkçe 2005	25	1031	13.09	30,3	5,50	0,846	0,524	0.100	-0,947

Her bir maddenin ortalama güçlüğü hesaplanıp, bu güçlüklerin ortalamaları alınıp, testin ortalama güçlüğü hesaplanmıştır. Her iki testin ortalama madde güçlükleri arasındaki farka, bağımsız

gruplar için oran testi ile bakılmıştır. 2003 yılı ve 2005 yılı OKS Türkçe testlerinin güçlükleri arasında fark olup olmadığına ait bilgiler Tablo-3’de verilmiştir.

Tablo 3: 2003 Yılı ve 2005 Yılı Şans Başarısından Arındırılmamış Test Puanlarının Ortalama Güçlük Düzeylerinin Karşılaştırılması

Alt Testler	\bar{p}	t	P
Türkçe 2003	0,539	0,68	0,363
Türkçe 2005	0,524		

Tablo-3 incelendiğinde, testlerin ortalama güçlükleri arasında 0,05 düzeyinde manidar bir fark olmadığı görülmüştür. Eşitlenmek istenen testlerin, ortalama güçlüklerinin aynı olması şartı sağlanmaktadır.

Eşitlenmek istenen testlerin eşit güvenilirlikte olup olmadığını belirlemek için, her bir alt testin KR-20 güvenilirlik katsayıları hesaplanmıştır. Güvenirlik katsayıları iki korelasyon katsayısı olarak kabul edilmiş ve güvenilirlik katsayılarına Fischer’in Z dönüştürmesi yapılmıştır. İki güvenilirlik katsayısı arasında manidar bir fark olup olmadığı Fischer’in Z istatistiği ile test edilmiştir. Güvenirlik katsayıları arasındaki farkın test edilmesine ilişkin bulgular, Tablo-4’te verilmiştir.

Tablo 4: 2003 Yılı ve 2005 Yılı Şans Başarısından Arındırılmamış Test Puanlarının Güvenirliklerinin Karşılaştırılması

Alt Testler	KR-20	Zr	Z
Türkçe 2003	0,853	1,242	0,569
Türkçe 2005	0,846	1,267	

Tablo-4 incelendiğinde, testlerin güvenilirlikleri arasında 0,05 düzeyinde istatistiksel olarak manidar bir fark olmadığı gözlenmiştir. Bu sonuca dayanarak eşitlenmek istenen puanların aynı güvenilirlik düzeyine sahip olduğu söylenebilir.

Diğer eşitleme şartları ise, iki testin ortalamaları ve varyanslarının eşit olmasıdır. Alt testlerin ortalamaları arasındaki farka bağımsız gruplar için t testi ile bakılmıştır. İki testin varyansları arasındaki farka F testi ile bakılmıştır. Bulunan sonuçlar Tablo-5’de verilmiştir.

Tablo 5: 2003 Yılı ve 2005 Yılı Şans Başarısından Arındırılmamış Test Puanlarının Ortalamalarının ve Varyanslarının Karşılaştırılması

Alt Testler	\bar{X}	t	p	S^2_x	F	P
Türkçe 2003	13.476	1,553	0,121	32,773	1,08	0,202
Türkçe 2005	13.092			30,268		

Tablo-5 incelendiğinde, testlerin ortalamaları ve varyansları arasında manidar bir fark olmadığı gözlenmiştir.

Tüm bu test sonuçlarına dayanarak, şans başarısından arındırılmayan test puanları için eşitleme şartlarının sağlandığı söylenebilir. Bu işlemlerden sonra, şans başarısından arındırılan ham puanlar için eşitleme şartlarının sağlanıp sağlanmadığı kontrol edilmiştir. İlk olarak şanstın arındırılmış test puanları arasındaki korelasyon 0,78 olarak hesaplanmıştır. Bu korelasyonun oldukça yüksek olduğu söylenebilir. Şans başarısından arındırılan ham puanlardan elde edilen bazı betimsel istatistikler Tablo 6’da verilmiştir.

Tablo 6: Şans Başarisından Arındırılmış 2003 ve 2005 OKS Türkçe Alt Test Puanlarına Ait Bazı Betimsel İstatistikler

Alt Testler	K	N	\bar{X}	S^2x	Sx	KR-20	Ortalama Güçlük	Çarpıklık katsayıları	Basıklık Katsayıları
Türkçe 2003	25	1031	9,916	57,30852	7,57	0,9338	0.397	0,021	-1,063
Türkçe 2005	25	1031	9,359	53,43419	7,31	0,9337	0.374	0.064	-0,950

Tablo-6 incelendiğinde, şanstın arındırılmış puanlara ait ortalama ve ortalama güçlüğü düşüğü; güvenilirlik ve varyansın yükseldiği gözlenmektedir. Bu durum literatürle benzerlik göstermektedir (Telli, 1993; Şahhüseyinoğlu, 1998; Araz, 2001). 2003 ve 2005 OKS Türkçe alt testlerinin ortalama güçlük indeksleri arasında manidar bir fark olup olmadığı, iki oran arasındaki farkın test edilmesi suretiyle irdelenmiştir. Sonuçlar Tablo-7’de verilmiştir.

Tablo 7 : 2003 Yılı ve 2005 Yılı Şans Başarisından Arındırılmış Test Puanlarının Ortalama Güçlük Düzeylerinin Karşılaştırılması

Alt Testler	\bar{p}	t	p
Türkçe 2003	0,397	1,095	0,283
Türkçe 2005	0,374		

Tablo-7 incelendiğinde, 2003 ve 2005 yıllarına ait testlerin ortalama güçlükleri arasında istatistiksel olarak anlamlı bir fark olmadığı gözlenmiştir. Eşitlenmek istenen testlerin eşit güvenilirlikte olup olmadığına ilişkin sonuçlar Tablo-8’de verilmiştir.

Tablo 8: 2003 Yılı ve 2005 Yılı Şans Başarisından Arındırılmış Test Puanlarının Güvenirliklerinin Karşılaştırılması

Alt Testler	KR-20	Zr	Z
Türkçe 2003	0,9338	1,687	
Türkçe 2005	0,9337	1,6865	0,0113

Tablo-8 incelendiğinde, testlerin güvenilirlikleri arasında istatistiksel olarak manidar bir fark olmadığı görülmüştür. Diğer eşitleme şartları ise, iki testin ortalamaları ve varyanslarının eşit olmasıdır. 2003 ve 2005 OKS Türkçe testlerinin ortalamaları arasındaki farka bağımsız gruplar için t testi ile; varyanslar arasındaki farka F testi ile bakılmıştır. Sonuçlar Tablo-9’da verilmiştir.

Tablo 9: 2003 Yılı ve 2005 Yılı Şans Başarisından Arındırılmış Test Puanlarının Ortalamalarının ve Varyanslarının Karşılaştırılması

Alt Testler	\bar{X}	t	p	S^2x	F	p
Türkçe 2003	9,916	1,701	0,0894	57,30852	1,07	0,261
Türkçe 2005	9,359			53,43419		

Tablo-9 incelendiğinde, testlerin ortalamaları ve varyansları arasında manidar bir fark olmadığı gözlenmiştir.

II. Aşama: Verilerin analizinin ikinci aşamasında, doğrusal ve eşit yüzdelliği eşitleme çalışması yapılmıştır.

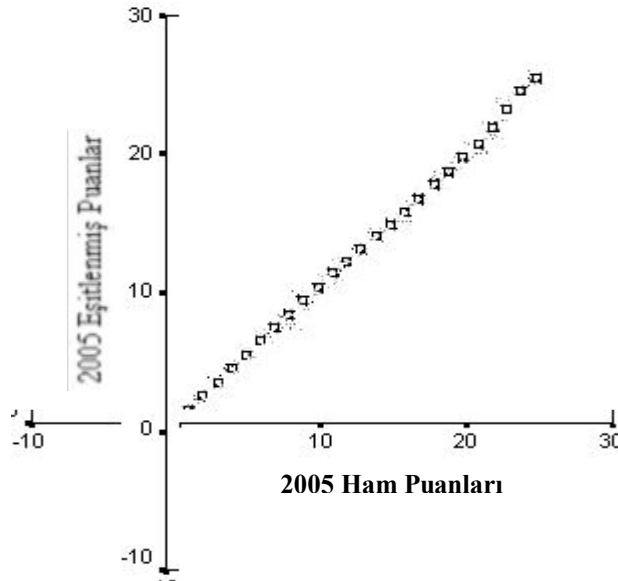
Doğrusal eşitleme için Crocker ve Algina'nın (1986) "Düzenek I" yöntemi kullanılarak 2003 OKS Türkçe testi puanları, 2005 Türkçe testi puanlarına eşitlenmiştir. Eşitlenmiş puanları elde etmek için (1) no'lu eşitlikten verilen denklemden yararlanılmıştır. Doğrusal eşitleme yönteminin, iki testin puan dağılımları benzer olduğu durumlarda kullanılması uygundur (Thorndike, 1982; Skaggs and Lissitz, 1986; Crocker ve Algina, 1986; Woldbeck, 1998). İki dağılımın benzer olup olmadığı, her iki testin puan dağılımına ait çarpıklık ve basıklık katsayılarına bakılarak incelenmiştir. Tablo-3 ve Tablo7'de verilen şanstın arındırılmamış ve arındırılmamış puanlara ait çarpıklık katsayıları incelendiğinde her iki katsayının da birbirine yakın olduğu ve sıfıra yaklaştığı gözlenmektedir. Bu bulgulara göre dağılımların birbirine benzer olduğu söylenebilir. İki teste ait ham puanlar ve frekans dağılımları ve toplamı frekansları bulunmuştur. Toplamı frekanslar bulunduktan sonra, iki testin ham puanlarının yüzdelik sıraları hesaplanmıştır. Ham puanların yüzdelik sıraları aynı sırada olmadığı için Livingston (2004) (3) no'lu eşitlikte verilen öteleme formülünden yararlanılarak uygulanmıştır.

III. Aşama: Son aşamada ise eşitleme hatası hesaplanmıştır. Eşitleme yöntemlerinin uygunluğunun ölçüsü olarak (4) no'lu eşitlikte verilen ağırlıklandırılmış hata kareleri ortalaması kullanılmıştır.

3. BULGULAR VE YORUM

3.1. Şans başarısından arındırılmamış 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, eşit yüzdelikli eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlara ilişkin bulgular.

Bu alt problem için ilk olarak, şans başarısından arındırılmamış 2003 yılı Türkçe test puanları, 2005 yılı puanlarına Crocker ve Algina'nın (1986) tek grup düzeneğine dayalı, eşit yüzdelikli eşitleme yöntemi ile eşitlenmiştir. 2005 yılı ham puanları ve bu puanlara karşılık gelen eşitlenmiş puanlara ait grafik Şekil 2'de verilmiştir.



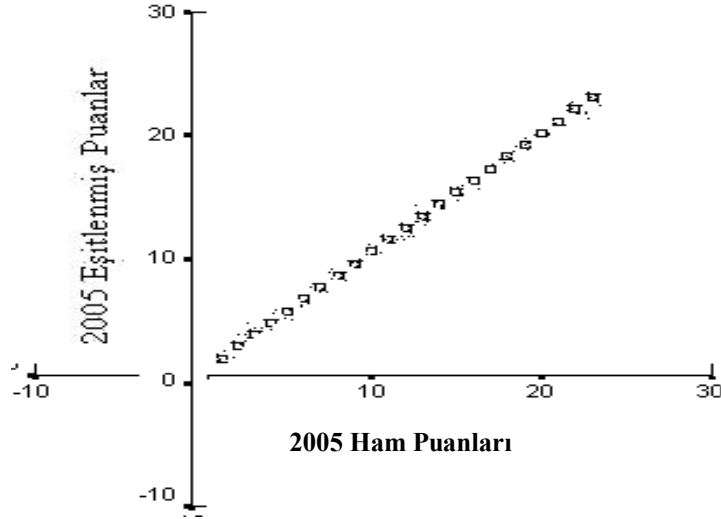
Şekil 2: Şans başarısından arındırılmamış 2003 yılı test puanlarının 2005 yılı test puanlarına eşit yüzdelikli eşitleme yöntemi ile eşitlenmesi

Şekil 2 incelendiğinde, eşitlenmiş puanlar ile 2005 yılı ham puanlar arasında doğrusal bir ilişki olduğu gözlenmektedir. Bu ilişki $Y=1,024X-0,007$ denklemi ile ifade edilebilir. Grafik incelendiğinde doğrunun eğiminin 1'e çok yakın olduğu gözlenmektedir.

Her iki teste ait ham puanlar nadiren aynı yüzdelik sıraya denk geldiği için Livingston (2004) tarafından önerilen öteleme formülü uygulanmıştır.

3.2. Şans başarısından arındırılmamış 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, doğrusal eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlara ilişkin bulgular.

2003 yılı puanları, 2005 yılı puanlarına Crocker ve Algina'nın (1986) tek grup düzeneği kullanılarak, doğrusal eşitleme yöntemi ile eşitlenmiştir. Ham puanlar ve eşitlenmiş puanlara ait grafik Şekil 3'te verilmiştir.



Şekil 3: Şans başarısından arındırılmamış 2003 yılı test puanlarının 2005 yılı test puanlarına doğrusal eşitleme yöntemi ile eşitlenmesi

Şekil 3'teki grafiğe bakıldığında ham puanlar ve eşitlenmiş puanlar arasındaki ilişkinin doğrusal olduğu görülür. Bu ilişki matematiksel olarak $Y=1.041x-0.147$ denklemi ile ifade edilebilir.

3.3. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılmadan eşitlenmesinde eşit yüzdelikli eşitleme ve doğrusal eşitleme yöntemlerinden hangisi daha uygundur?

Bu alt problemde bu iki yönteme ait ağırlıklandırılmış hata kareleri ortalaması bulunmuş ve ağırlıklandırılmış hata kareleri ortalaması en küçük olan yöntem, şans başarısından arındırılmadan yapılan eşitleme için en uygun yöntem olarak kabul edilmiştir. Şans başarısından arındırılmamış 2003 yılı puanlarının 2005 yılı puanlarına eşitlenmesine ait ağırlıklandırılmış hata kareleri ortalamaları Tablo-10'da verilmiştir.

Tablo 10: Şans Başarısından Arındırılmamış Puanların Doğrusal ve Eşit Yüzdelikli Eşitleme Yöntemleri Kullanılarak Elde Edilen Ağırlıklandırılmış Hata Kareleri Ortalamaları (AHKO)

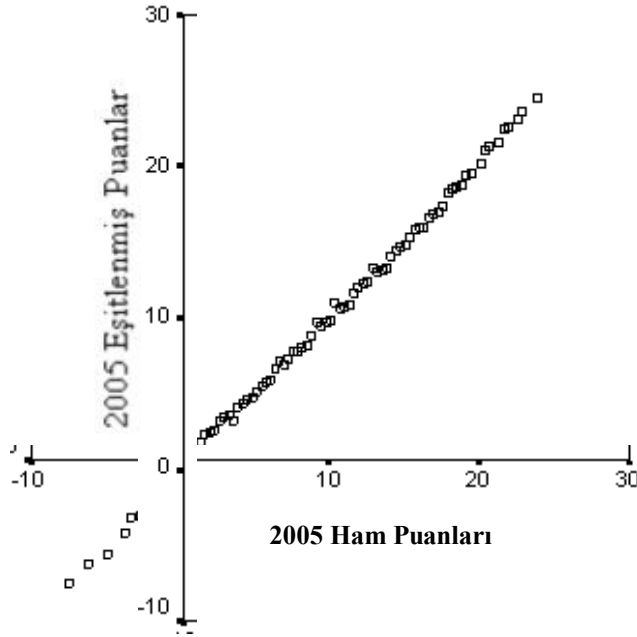
A.H.K.O.	Eşitleme Yöntemleri	
	Eşit Yüzdelikli Eşitleme	Doğrusal Eşitleme
	0,007	0,006

Şanstan arındırılmamış 2003 ve 2005 OKS Türkçe alt test puanlarının doğrusal eşitleme yöntemi kullanılması sonucu elde edilen hata kareleri ortalaması (eşitleme hatası) 0,006 olarak bulunmuştur. Şanstan arındırılmamış ham puanların, eşit yüzdelikli eşitleme yöntemi kullanılarak eşitlenmesinde ise hata kareleri ortalaması 0,007 olarak bulunmuştur. Skaggs ve Lissitz (1986), hata kareleri ortalamalarının manidarlığının belirlenmesinde istatistiksel bir test olmadığını ve bu hataya ilişkin değerlerin 0,05 veya daha büyük olması durumunda pratikte anlamlı olabileceğini belirtir. Her iki

yöntem için de hata karelerinin ortalamalarının 0,05'ten küçük ve birbirine yakın değerler olduğu görülmektedir. Eşitleme hatalarının yakın değerler almasının, dağılımların birbirine benzer olmasından kaynaklandığı düşünülmektedir (Felan, 2002). Doğrusal eşitlemeye ait hata daha küçük olduğu için şans başarısından arındırılmadan yapılan eşitlemede, doğrusal eşitlemenin daha uygun bir yöntem olduğu söylenebilir. Bu bulgu Budescu'nun (1987) bulguları ile tutarlılık göstermektedir. Bu durumun ortaya çıkmasında, eşitleme koşullarının karşılanmasının etkili olduğu düşünülmektedir.

3.4. Şans başarısından arındırılan 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, eşit yüzdelikli eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlara ilişkin bulgular.

2003 yılı şans başarısından arındırılmış ham puanların, 2005 yılı şans başarısından arındırılmış ham puanlara eşit yüzdelikli eşitleme yöntemiyle eşitlenmiştir. Ham puanlar ve eşitlenmiş puanlara ait grafik Şekil 4'te verilmiştir.



Şekil 4: Şans başarısından arındırılmış 2003 yılı test puanlarının 2005 yılı test puanlarına eşit yüzdelikli eşitleme yöntemi ile eşitlenmesi

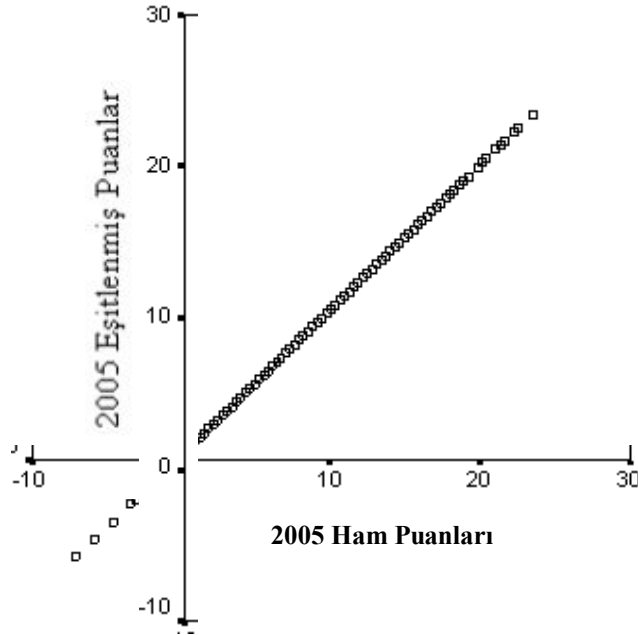
Şekil 4 incelendiğinde, eşitlenmiş puanlar ile ham puanlar arasındaki ilişkinin doğrusal olduğu görülmektedir. Bu ilişki matematiksel olarak $Y=0,977X+0,002$ denklemi ile ifade edilir.

2003 yılı OKS Türkçe alt testi şanstın arındırılmış puanları, 2005 yılı puanlarına eşit yüzdelikli eşitleme yöntemi ile eşitlenmiştir. Şans başarısından arındırılmadan yapılan eşit yüzdelikli eşitleme yönteminde olduğu gibi (3) no'lu eşitlikten yararlanılmıştır.

3.5. Şans başarısından arındırılan 2003 ve 2005 yılları OKS Türkçe testi ham puanlarının, doğrusal eşitleme yöntemi kullanılması sonucu elde edilen eşitlenmiş puanlara ilişkin bulgular.

Doğrusal eşitleme yöntemi için öncelikle iki testin şans başarısından arındırılmış puanlarının ortalaması ve standart sapması bulunmuş ve 2003 puanları, 2005 puanlarına eşitlenmiştir.

Ham puanlar ve eşitlenmiş puanlara ait grafik Şekil 5'te verilmiştir.



Şekil 5 Şans başarısından arındırılmış 2003 yılı test puanlarının 2005 yılı test puanlarına doğrusal eşitleme yöntemi ile eşitlenmesi

Şekil-5'te görüldüğü gibi, puanlar arasındaki fark uçlarda fazla ve orta noktalarda hemen hemen aynıdır. Eşitlenmiş puanlar ile ham puanlar arasındaki ilişki $Y=0,932X+0,113$ denklemi ile ifade edilebilir.

3.6. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılarak eşitlenmesinde eşit yüzdelli eşitleme ve doğrusal eşitleme yöntemlerinden hangisi daha uygundur?

Tablo-11'de şanstın arındırılmış puanların doğrusal ve eşit yüzdelli eşitleme yöntemine ait ağırlıklandırılmış hata kareleri ortalamaları verilmiştir.

Tablo 11: Şans Başarısından Arındırılmış Puanların Doğrusal ve Eşit Yüzdelli Eşitleme Yöntemleri Kullanılarak Elde Edilen Ağırlıklandırılmış Hata Kareleri Ortalamaları (A.H.K.O)

	Eşitleme Yöntemleri	
	Eşit Yüzdelli Eşitleme	Doğrusal Eşitleme
A.H.K.O.	0,0054	0,0096

Tablo-11 incelendiğinde, şans başarısından arındırılan ham puanların eşit yüzdelli eşitleme kullanılarak eşitlenmesine yönelik bulunan ağırlıklandırılmış hata kareleri ortalaması 0,0054; doğrusal eşitleme yöntemi kullanılarak hesaplanan eşitlenmiş puanlar ile ham puanlara ilişkin ağırlıklandırılmış hata kareleri ortalaması 0,0096 olarak görülmektedir.

Her iki yöntemle ilişkin eşitleme hatasının, 0,05'ten küçük olduğu görülmüştür. Eşit yüzdelli eşitleme yöntemine ait ağırlıklandırılmış hata kareleri ortalaması daha küçük olduğu için, eşit yüzdelli eşitleme yönteminin daha uygun olduğu söylenebilir.

3.7. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındırılarak ve arındırılmayarak eşitlenmesinde kullanılan doğrusal eşitleme yöntemlerinden hangisi daha uygundur?

Tablo 12: Şans Başarisından Arındirilmamis ve Arindirilmis Puanların Dogrusal Eşitleme Yöntemi ile Eşitlenmesine Yönelik Ağırlıklandırılmış Hata Kareleri Ortalaması

AHKO	Doğrusal Eşitleme Yöntemi
Şanstan Arındirilmamis	0,006
Şanstan Arındirilmis	0,0096

Tablo-12 incelendiğinde şanstan arındirilmamis test puanlarının dogrusal eşitleme ile eşitlenmesine ait ağırlıklandırılmış hata kareleri ortalamasının daha küçük olduğu gözlenmektedir. Daha önce eşitleme koşullarına ilişkin yapılan analizler sonucunda, şans başarısından arındirılan puanlara ait ortalama ve varyanslar arasındaki farkın arttığı gözlenmiş ancak bu farkın manidar olmadığı belirlenmiştir. Ortalama ve varyanslar arası farkın artmasının, şans başarısından arındirilmiş puanların dogrusal eşitlemesine yönelik eşitleme hatasını arttırabileceği ileri sürülebilir.

3.8. 2003 ve 2005 yılı OKS Türkçe alt testlerinin ham puanlarının şans başarısından arındirılarak ve arındirilmayarak eşitlenmesinde kullanılan eşit yüzdellikli eşitleme yöntemlerinden hangisi daha uygundur?

Tablo 13: Şans Başarisından Arındirilmamis ve Arindirilmis Puanların Eşit Yüzdellikli Eşitleme Yöntemi ile Eşitlenmesine Yönelik Ağırlıklandırılmış Hata Kareleri Ortalaması

AHKO	Eşit Yüzdellikli Eşitleme Yöntemi
Şanstan Arındirilmamis	0,007
Şanstan Arındirilmis	0,0054

Tablo-13 incelendiğinde, şanstan arındirilmiş puanların eşit yüzdellikli eşitlenmesi sonucu hesaplanan ağırlıklandırılmış hata kareleri ortalaması daha küçük olduğu görülmüştür. Bu durumun nedeninin, eşit yüzdellikli eşitleme yöntemi için gerekli olan sık puan aralığından kaynaklandığı düşünülebilir. Eşit yüzdellikli eşitleme için puan aralığının sık ve gergin olması ile ham puan dağılımının diğer puan dağılımına denk gelmesi sağlanır (Angoff, 1971; Woldbeck, 1998). Şanstan arındirilmamis eşitlemede puan aralığı 1 iken, puanlar şanstan arındirıldığında puan aralığı 0,33 olmuştur. Bu durumun, eşit yüzdellikli eşitleme yöntemine ait hatanın küçülmesine neden olduğu düşünülebilir.

4. SONUÇ VE ÖNERİLER

Bu çalışmada, 2003 ve 2005 OKS Türkçe alt testlerine ait ham puanların şans başarısından arındirılmadan ve arındirılarak eşitlenmesinde dogrusal ve eşit yüzdellikli eşitleme yöntemlerinden hangisinin daha uygun olduğu incelemiştir. Araştırmaya ait sonuç ve öneriler aşağıda verilmiştir.

4.1. Sonuçlar:

1. Şans başarısından arındirilmamis 2003 yılı OKS Türkçe alt test puanlarının, 2005 yılı puanlarına dogrusal ve eşit yüzdellikli yöntem kullanılarak eşitlenmesinde, dogrusal eşitlemeye ait eşitleme hatasının daha küçük olduğu görülmüştür.
2. Şans başarısından arındirilmiş 2003 yılı OKS Türkçe alt test puanlarının, 2005 yılı puanlarına dogrusal ve eşit yüzdellikli yöntem kullanılarak eşitlenmesinde, eşit yüzdellikli eşitlemeye ait eşitleme hatasının daha küçük olduğu gözlenmiştir.
3. Şanstan arındirilmiş puanların eşitlenmesinde, ağırlıklandırılmış hata kareleri ortalaması en küçük olan yöntem, eşit yüzdellikli eşitleme yöntemi olarak bulunmuştur.

4. Ham puanlar şanstın arındırıldığında, doğrusal eşitlemeye ait hatanın arttığı görülmüştür.
5. Ham puanların şanstın arındırılması sonucu, puan ranjının genişlediği ve puan aralıklarının daraldığı gözlenmiştir. Bu durumun, eşit yüzdelikli eşitlemeye ait hatanın küçülmesine neden olduğu düşünülmektedir.
6. İki puan dağılımının birbirine benzer olması sonucunda, eşitlenmiş puanların benzer ve eşitleme hatalarının birbirlerine oldukça yakın olduğu görülmüştür.
7. Tek grup düzeneğinde, birbirine eşitlenecek olan iki veya daha fazla test aynı gruba uygulanır. Uygulaması oldukça basit ve pratik bir yöntemdir. Aynı grubun her iki formuda alması bu düzeneği güçlü kılan bir özelliktir ve bu sebeple de bu düzenek aracılığıyla elde edilen eşitleme kesinliği diğer düzeneklere göre daha yüksek, eşitleme hatası ise daha düşüktür (Livingston, 2004; Felan 2002; Zhu, 1998; Dorans, 1990; Crocker ve Algina 1986). Bu bulgulara dayanarak, araştırmada tek grup düzeneğinin kullanılması, eşitleme hatasına ilişkin değerin oldukça küçük çıkmasına neden olduğu söylenebilir.

4.2. Öneriler:

Öneriler aşağıda, araştırmının devam ettirilmesine ve kullanım alanı bulmasına yönelik olmak üzere iki kısımda verilmiştir.

Araştırmının devam ettirilmesine yönelik öneriler:

1. Bu çalışmada tek grup düzeneği kullanılmıştır. Diğer düzenekler düzeneği kullanılarak araştırma yeniden yapılabilir. Böylece sonuçlar birbiriyle karşılaştırılabilir.
2. Araştırma, madde sayısı daha fazla olan testler üzerinde, daha büyük gruplarda tekrarlanabilir.
3. Farklı şans başarısından arındırma teknikleri kullanılarak, şans başarısının test eşitlemeye etkisi incelenebilir.
4. Araştırmada klasik yöntemlerle eşitleme yapılmıştır, IRT eşitleme yöntemleri denenebilir.
5. Daha önce uygulanmış iki farklı forma ait veriler üzerinde, ankor madde düzeneği kullanılarak, şans başarısının test eşitlemeye etkisi incelenebilir.
6. Çoktan seçmeli testlerde, seçenek sayısının test eşitlemeye etkisi, şans başarısı ile birlikte incelenebilir.

Araştırmının kullanım alanı bulmasına yönelik öneriler:

KPSS, ALES ve KPDS... gibi seçme amaçlı yapılan sınavlardan alınan puanlar birkaç yıl süreyle kullanılabilir. Fakat farklı zamanlarda bu sınavlardan alınan puanların aynı ölçek üzerinde yer almadıkça eşdeğer kabul edilmesi doğru değildir. Puanların aynı kabul edilebilmesi için bir formun birim sisteminin diğer formun birim sistemine dönüştürülmesi gerekir. Bu dönüşüm test eşitlemeyle sağlanır. Ülkemizde uygulanan bu sınavlardan bazılarında ham puanlar hesaplanırken şans başarısından arındırılır (KPSS). Bazı sınavlarda ise ham puanlar hesaplanırken şans başarısından arındırılmaz (KPDS). Bu çalışmanın bulguları doğrultusunda, şanstın arındırılmamış ham puanların eşitlenmesinde doğrusal eşitleme yöntemi, şans başarısından arındırılmış puanların eşitlenmesinde ise eşit yüzdelikli eşitleme yönteminin kullanılması önerilebilir.

KAYNAKLAR

- Angoff, W. H. (1971). "Scales, Norms and Equivalent Scores". Thorndike, R. L. (Ed.) *Educational Measurement*. Washington, D. C. American Council on Education.
- Araz, G. (2001). " Aynı Davranışı Ölçmeye Yönelik Kısa Cevaplı, Üç ve Beş Seçenekli Çoktan Seçmeli Testlerin Madde ve Test Özelliklerini Şans Başarısı ile Birlikte İncelenmesi" Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi SBE.
- Barnard, J. J. (1996). "In Search For Equity in Educational Measurement: Traditional Versus Modern Equating Methods". Paper presented at ASEESA's National Conference, Pretoria South Africa.
- Budescu, D. V. (1987). "Selecting an Equating Method: Linear or Equipercentile?". *Journal of Educational Statistics*. Vol.12s, No:1. 33-43.

- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. CBS Collage Publishing, New York.
- Çelen, Ü. (2002). “Şans Başarısı için Düzeltme Formülü Kullanılacağına İlişkin Yönergenin Testin Psikometrik Özelliklerine Etkisinin Araştırılması” Yayınlanmamış Yüksek Lisans Tezi, Ankara Üniversitesi SBE.
- Dorans, N. J. (1990). Equating methods and sampling design. *Applied Measurement In Education*. 3(1), 3-17.
- Dorans, N. J. (2000). “Research Notes: Distinctions Among Classes of Linkages.” *The College Board, Office of Research and Development*.
- Felan, G. D. (2002). “Test Equating: Mean, Linear, Equipercentile and Item Response Theory”. Paper presented at the *Annual Meeting of the Southwest Educational Research Association*, Austin.
- Hambelton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer Academic Publishers Group, Boston
- Harris, D. J. (2003). “Equating the Multistate Bar Examination” Reprinted with permission of the *National Conference of Bar Examiners*. Vol:12, No:3. 12-18.
- Gulliksen, H. (1967). *Theory of Mental Tests*. New York: John Wiley & Sons.
- Kelecioğlu, H. (1993). “Öğrenci Seçme Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalışma”. Yayınlanmamış Doktora Tezi, H.Ü. Sosyal Bilimler Enstitüsü.
- Kelecioğlu, H. (1995). “Ortak Testlerin Eşitleme Hatasına Etkisi”. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 11s,121-127.
- Kolen, M. J. (1988). “An NCME Instructional Module on Traditional Equating Methodology”. *Educational Measurement: Issues And Practice*. v7, n4:29-36
- Livingston, S. A. (2004). “Equating Test Scores (Without IRT)”. *Educational Testing Service*.
- Másse, L. C.; Allen, Diane; Wilson, Mark; ve Williams, Geoffrey (2006). “Introducing equating methodologies to compare test scores from two different self-regulation scales”. *Health Education Research* 21, 110-120.
- Skagg, G., & Lissitz R. W. (1986). “An Exploration of the Robustness of Four Test Equating Models” *Applied Psychological Measurement*. 10, 303-317.
- Şahhüseyinoğlu, D. (1998). “Sayısal Yetenek Testlerinde Seçenek Sayısının Test ve Madde İstatistikleri Üzerindeki Etkisinin Şans Başarısı ile Birlikte İncelenmesi” Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi SBE.
- Şahhüseyinoğlu, D. (2005). “İngilizce Yeterlik Sınavı Puanlarının Üç Farklı Eşitleme yöntemine Göre Karşılaştırılması” Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi SBE.
- Tanguma, J. (2000). “Equating Test Scores Using Linear Method”. Paper presented at the *Annual Meeting of the Southwest Educational Research Association*, Dallas.
- Telli, A. (1993). “Şans Başarısının Madde Türlerindeki Madde ve Test İstatistiklerine Etkisi”. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi SBE.
- Thorndike, R. L. (1982). *Aplied Psychometrics*. Houghton Mifflin Company, Boston.
- Tsai, T. H. (1997). “Estimating Minumum Sample Sizes in Seçkisiz Groups Equating”. Paper presented at the *Annual Meeting of the National Council on measurement in Education*, Chicago.
- Turgut, F. (1971). *Şans Başarısının Test Puanlarına Etkisi*, ODTÜ Yayınları, Ankara.
- Turgut, F. (1979). “Ölçekleme, Norm Hazırlama, Normları Eşitleme ve Ayarlama”. Test Geliştirme Teknikleri Basılmamış Ders Notları, Ankara.
- Turgut, F. (1988). *Eğitimde Ölçme Değerlendirme*, Saydam Yayıncılık, Ankara.
- Woldbeck, T. (1998). “Basic Concepts in modern Methods of Test Equating”. Paper presented at the *Annual Meeting of the Southwest Psychological Association*, New Orleans.
- Zhu, W. (1998). “Test Equating: What, why, how? Research”. *Quarterly for the Exercise and Sport*, 69, 11-23.

Extended Abstract

In our country, in order to select students and staff for various institutions, tests such as Student Selection Examination (OSS), High Schools Student Selection and Placement Examination (OKS), Public Personnel Selection Examination (KPSS) etc. are applied every year. Though the behaviours tested every year remain the same, different forms are applied on individuals to ensure test security. The method developed for comparing the scores obtained from different forms is called test

equation. Test equation method controls whether any of the various forms of tests measuring the same structure, provides an advantage for the group taking it, or not.

The objective of this study is to find out whether chance success effects test equating, or not. Single group design, and linear equating and equipercentile equating from classical equating methodologies were used in the study. Furthermore, it's aimed to obtain equated scores using these equation methods, to establish a relationship between equated scores and raw scores, to find substitute scores in parallel to this correlation, and to suggest the suitable equation method. No study examining the effect of chance success on test equating could be found in the literature scan. Therefore, this study is considered to contribute to the studies in the field of test equation.

The data of the study was collected from 1031 8th grade students from various primary schools in Mersin province, in May of 2005-2006 Academic Year, by means of Turkish language sub-tests of OKS 2003 and 2005. The scores of Turkish language sub-test of 2003 OKS were equated to those of 2005 OKS using the linear equating and equipercentile equating methods, with and without the chance success. For elimination of chance success in the scores, the correction formula in which some of the wrong answers are subtracted from the number of the correct answers was employed. The corrected test scores (CTS) were calculated by this formula:

$$CTS = R - \frac{W}{k}$$

C: Total number of correct answers

W: Total number of wrong answers

k: number of options

The data analysis was performed at three stages. At the first stage, it was checked if equating conditions are fulfilled. It was checked whether two tests to be equated measure the same structure or not, by combining the individual – item matrix of two tests and conducting factor analysis. It was established that they both measure the same structure. Another equation condition is that both tests are single dimensional. To ensure this, factor analysis of the test scores were carried out separately, and it was identified that they both are single-dimensional. The correlation between the tests to be equated should be high; the correlation between the two tests scores was calculated with and without the chance success and the correlations were high in both cases. The mean scores, mean difficulties, reliabilities and variances of both tests with chance success were compared, and no statistically significant difference could be identified. This demonstrates that the equating conditions are fulfilled for scores with chance success. The same statistics were calculated for the scores without chance success, and it was established that mean scores, mean difficulties, reliabilities and variances of both tests do not display a statistically significant difference, hence fulfilling the equating conditions.

At the second stage of data analysis, the scores with the chance success were equated using linear and equipercentile equating methods to obtain equated scores. The same methods were applied for scores without chance success of Turkish language sub-tests of 2003 and 2005 OKS, and equated scores were obtained. At the third stage, weighted mean square error was calculated as conformity scale of equation methods. The method with the lowest weighted mean square error was suggested as the most suitable equating method.

At the findings part of the study, the scores of Turkish language sub-test of 2003 OKS were equated to those of 2005 OKS using the linear equating and equipercentile equating methods, with and without the chance success to find equated scores; and weighted mean square errors were calculated out of these equated scores. Consequently, it was found out that equation error of linear equating is smaller when Turkish language sub-test scores of 2003 OKS are equated to 2005 scores with chance success using linear and equipercentile equating methods. As for the equating of scores without chance success, it was observed that equation error of equipercentile equating is smaller. When raw score were calculated without chance success, it was observed that the equating error of linear equating increased while that of equipercentile equating decreased. As a result of single group design being employed in the study and similarity of the two distributions, the values of the equation errors were small and quite close to each other.

The suggestions part of study was gathered under two headings. Those are suggestions for carrying on and using the study. In order to carry on the study to a further stage the following research

may be conducted: the study may be repeated using balanced group design, may be repeated for tests with more items on bigger groups, or the effect of chance success on test equating may be investigated using different techniques for eliminating chance success. In this study, the equation was performed through classical methods; IRT equation methods might be tried. Chance success effect on test equating may be examined using anchor design on the data of two different forms applied earlier; for multiple choice tests, the effect of the number of choices on test equating may be examined together with chance success.

The scores obtained from selection examinations such as KPSS, KPDS and ALES are being used for several years in our country. Unless the scores of different tests at different times are on the same scale, it's not accurate to accept them to be equivalents. Test equation would make it possible to place scores of different forms onto the same scale. In parallel to the findings of this study, the linear equating method for the equation of scores with chance success and equipercentile equating method for equation of scores without chance success could be recommended.